

Customer Questionnaire

Introduction

The questions in this document are intended to aid in the planning for your storage and retrieval of data from the Fermilab archive. You may not yet know the answers to some of the questions so please do your best and if you could, note any uncertainties.

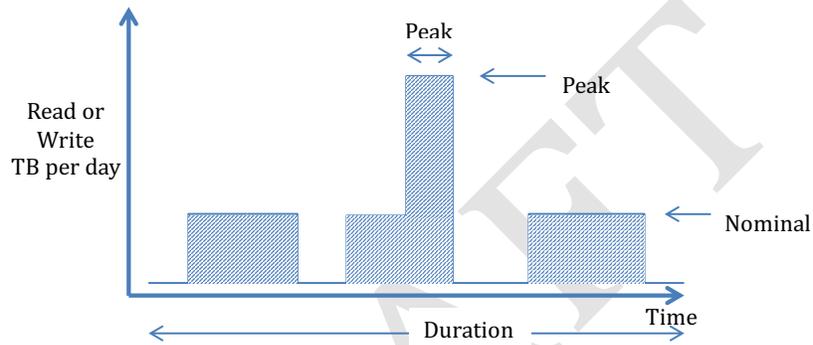
Definitions

Please refer to Figure 1.

- All data and file sizes should be expressed in Gigabytes (1 GB = 10^9 bytes) or Terabytes (1 TB = 10^{12} bytes)
- Bandwidth in MB/s or GB/s (Mega or Giga bytes per second)
- Nominal data rate is the typical data rate while you are actively writing, excluding much larger peak activity.
- Peak data rate is highest data rate above the average that may occur because of special activity.
- Total TB is the integral of the data written over a period of time
- Duty factor is Total / (Average * Duration)
- Peak duration – the amount of time you expect peak activity to occur
- Periodicity – any periodic structure in data rate (e.g. peak storages in October, nominal the rest of the year)
- Writer: Application writing to storage
- Reader: Application reading from storage
- Client node: a node that either readers or writers are running on
- Client Center: Distinct client network location
- HSM: Hierarchical Storage Management. Fermilab storage is and HSM consisting of tape storage with a front-end disk cache. Your data will be transferred to/from the front-end disk cache. The HSM takes care of moving this data on/off tape from/to disk.
- Enstore: Tape management component of the Fermilab HSM
- dCache: Front-end disk cache storage for the Fermilab HSM
- Pre-staging to disk: Copying a set of files from tape to disk prior to accessing them. Pre-staging can optimize the retrieval of files from tape to disk prior to their retrieval by the client.
- Protected Personally Identifiable Information (PII): Data such as credit card number, SSN, medical records, that if made public, can cause harm to an individual:

http://www.fnal.gov/directorate/Directors_Policy/personally_identifiable_info.shtml

- Custodial contact: Customer contact responsible for making decisions on customer user authorization, access, data origination, and data deletion. This contact also will be put on outage notification lists and is responsible for communicating information to customer users about proper usage, service outages, and other service related information.



Integral gives Total TB
Duty factor = Total / (Nominal * Duration)

Figure 1.

Archiving (Write) Requirements

Requirement	Estimated Value	Comment
Nominal data rate (MB/s)		
Peak data rate (MB/s)		
Maximum Peak Duration (hours or days)		
Duty factor (%)		
Year 1 Storage (TB)		
Year 2 Storage (TB)		
Year 3 Storage (TB)		
Year 4 Storage (TB)		
Year 5 Storage (TB)		
Eventual total Storage (TB)		
Periodicity in storage activity (Y/N)? (If Y explain)		
Number of simultaneous writers		
Number of client nodes or client nodes per data center		
Number of client data centers or locations		
File size distribution (GB),		

Requirement	Estimated Value	Comment
minimum, maximum and average		
Retention period (yrs)		
Regulatory or other Restriction/Requirements for retention (Y/N). If Y explain.		

Retrieval (Read) Requirements

Requirement	Estimated Value/ Response	Comment
Average data retrieval rate		
Peak data retrieval rate (MB/s)		
Peak data retrieval duration (hours or days)		
Peak data retrieval size (TB)		For disk caching/pre-staging
Retrieval duty factor (%)		
Disk Retention (hours or days)		Data will normally be pre-staged from tape to disk for efficiency purposes.
Retrieval activity periodicity (Y/N)? (If Y explain)		
Number of simultaneous readers		
Number of client nodes or client nodes per data center		
Number of client data centers or locations		

Transfer Protocols Requirements

Do you have a preference of transfer protocols?

Protocol	Response	Comment
GSI FTP (Y/N)		
WEBDav (Y/N)		
SRM (Y/N)		
Globus Online (Y/N)		
Other (specify)		
Special security Requirement/preference (Y/N) (If Y, specify)		Most of the transfer protocols rely on authentication based on X.509 certificates.

Usability Requirements

Requirement	Response	Comment
What Metadata Manipulations do you require?		Examples: Deletions, rename, listing, mass deletions, etc
What types of monitoring data do you require?		Examples: Total storage, YTD, bandwidth usage, file listings, etc., billing information, etc.
How is your data organized?		

Formatted: French

Formatted: French

Security and Privacy Requirements

Requirement	Response	Comment
Do you have specific data privacy requirements (Y/N)? If Y please explain.	<u>no</u>	
Do you have specific metadata privacy requirements?	<u>no</u>	
Does any of your data contain Personally Identifiable Information (PII) (Y/N)? If Y please explain.	<u>no</u>	
Is any of your data encrypted (Y/N)?	<u>no</u>	
Does any of your data have any regulatory or other restrictions or requirements on security or privacy (Y/N)? If Y, please explain.	<u>no</u>	

Data Integrity Requirements

Requirement	Response	Comment
Is there any subset of your data that is critical and may warrant duplicate copies (Y/N)? If Y please elaborate.	<u>no</u>	
Do you have specific data loss requirements (Y/N)? If Y, what?	<u>no</u>	
Does any of your data need to follow any regulatory or other restrictions or requirements on integrity or validation (Y/N)? If Y, please explain	<u>no</u>	

Contacts

Please provide custodial, technical and business/financial contacts

Contact	Contact Name	email	Telephone number
Custodial	Albert Stebbins	stebbins@fnal.gov	x3663
Technical	Albert Stebbins	stebbins@fnal.gov	x3663
Business/financial	Peter Timbie	ptimbie@wisc.edu	(608) 890-2002

Formatted Table

Deleted: .

Other Requirements

Do you have any other requirements, comments or questions?